PNAS

Supporting Information for
Unequal resource division occurs in the absence of group division and identity

Eliane Deschrijver*[1,2,3]
Richard Ramsey[4, 5]

[1] School of Psychology, Faculty of Science, University of Sydney; A18 Manning Rd, Camperdown, NSW 2050, Australia.
[2] School of Psychological Sciences, Faculty of Medicine, Health and Human Sciences. Macquarie University, 16 University Avenue, NSW 2109, Macquarie Park, Australia.
[3] Department of Experimental Psychology, Ghent University, Henri Dunantlaan 4, Gent, 9000, Belgium
[4] Neural Control of Movement Laboratory, Department of Health Sciences and Technology and [5]Social Brain Sciences Laboratory, Department of Humanities, Social and Political Sciences, ETH Zürich, Gloriastrasse 37/39, 8092 Zürich, Switzerland.

*Corresponding author
Email: eliane.deschrijver@sydney.edu.au

This PDF file includes:

- Supporting text
- Tables S1 to S13
- Figures S1 to S8

Supporting Information Text

Materials And Methods

The experiments were programmed in JsPsych, and ran via the program JATOS on the server Mindprobe with participants recruited on the platform Prolific (www.prolific.co). Where possible, we implemented the standard instructions, experimental design and response matrices to introduce (the variations on) the minimal group paradigm (as described in [1]).

Data and Materials Availability

Across all seven experiments, the research questions, hypotheses, planned analyses, sample sizes and exclusion criteria were pre-registered before data collection started. The pre-registrations can be accessed at the following links (Exp. 1: osf.io/bzjxy, Exp. 2: osf.io/dxc5a, Exp. 3a: osf.io/4n8qz, Exp. 3b: osf.io/frb2t, Exp. 4: osf.io/q7hp9, Exp. 5: osf.io/snjg5, and Exp. 6: osf.io/snjg5). In addition, consistent with recent metascience proposals [2], the raw data, stimuli, and analysis code for each experiment are available online on the open science framework (osf.io/zd5ey). By doing so, others can pursue tests of alternative hypotheses, as well as more exploratory analyses and meta-analyses.

For Exp. 3b, we report one deviation from pre-registration. Our pre-registered regression models assumed a within-participant design for setting varying effects. However, Exp. 3b used a between-participant design, which means that it was not possible for the effect of task to be a between-participant effect. Therefore, in our main analysis of Exp. 3b below, we use the modelling approach that we pre-registered, except that the effect of group does not vary by participant because that is impossible in this design.

PART 1: Experiments Based on Social Psychology's Minimal Group Paradigms

Participants

We conducted Experiments 1, 2 and 3a with a sample of 238 participants each. Experiment 3b was conducted with a sample of 238 participants in each of the two participant groups. As we did not have initial data for the first 3 experiments, we used a benchmark of a small effect size to compute our sample size. The sample size of 238 participants was determined using an a-priori power-analysis for a matched pairs Wilcoxon signed-rank test [1] in G*Power, assuming a small effect size of 0.2 and 85% power. Eligibility criteria for participants are: no mental health conditions, use of the browser Chrome, from the UK, not participated in one of our earlier experiments, and answering 'yes' on Prolific's prescreening question about being comfortable with deception experiments. Recruitment was evenly distributed in terms of gender. To ensure quality control of our data, participants had a prior minimal approval rate of 95% and a minimum of 20 previous submissions. Participants were automatically excluded from further participation if they twice failed two attention checks ("Select a wrong answer to the question: How much is 8+8?"; and "My parents say I like aliens. Here, you need to indicate 'strongly agree'."). We excluded participants with incomplete datasets. Data collection for each experiment finished after completing a sample of 238 full datasets. We replaced participant data when they had experienced technical issues, revoked informed consent, or showed non-genuine or overly repetitive responding in line with our preregistered plan (Exp. 1: 8 datasets, Exp. 2: 6 datasets, Exp. 3a: 8 datasets, Exp. 3b: 3 datasets for the individual task and 4 datasets for the group task).

Design

In Experiments 1-3a, the participants completed two within-subjects blocks, with the individual block always preceding the group block to avoid priming them with a notion of groups. In Experiment 3b, the individual and the group task were assigned on a between-subjects basis. Each participant received 3 practice trials.

Experiment 1: Dots Estimations – Within-Subjects. After collecting informed consent and demographical data, participants were told that they were going to participate in a visual detection task. They were asked to estimate whether the number of blue dots was 'more' or 'less' than a certain reference number. Participants were instructed that we would use rewards to investigate other kinds of decisions as well, for understanding interpersonal differences.

In the 'individual' block, the participant was deceived in thinking that precisely one other participant had already completed the same trials, of which we had recorded the data. The participant was asked to assign points to this person, and the other person had ostensibly already done the same for them. These points would after completion of the entire data-collection be translated into a reward of up to £0.5 each. While some studies have shown that this may lead to reciprocity expectations, we still adhered here to this feature to stay as close as possible to the original minimal group design (yet, in experiments 4 to 6 we rule out any influence of this reciprocity). We explained that we would use response matrices to allocate the rewards to the other individual. The participant was told that raters were on average equally likely to choose the "more" or "less than the reference number" but that this was not related to their accuracy. This should avoid any influence on allocation strategies of an assumption that one or the other response option was more likely (as per [1]). Each trial started with a fixation cross presented for 500ms. The participant would then observe on the screen for 1500ms a cloud of small blue dots (with a radius of 4 pixels), which ranged from 20 to 70 dots in increments of 10, within a circle aperture with a diameter of 500 pixels. We then asked: "What do you think the number of dots in this trial was? The reference number is: <X>". This reference number <X> was always the exact number of dots on the screen. Dots estimation responses were recorded as: "Less than <X>" or "More than <X>". Participants then received a response matrix. We presented the following sentence presented underneath: "You estimated the number of dots as <more/less> than the reference number <X>. Think carefully." We emphasized that they were not assigning points to themselves, and that they could only choose points that belonged to the same set in the matrix (see supplementary figure 1). The participant then indicated their choice using the letter buttons A to M. A total of 6 types of matrices were presented once in the original format [1], and once in a flipped version. This resulted in 12 trials in total.

In the group block, we implemented the usual elements of the design of the minimal group paradigm. The participant would estimate the number of yellow dots on the screen. We deceived them in thinking that the next section of the experiment had been completed by *many* participants already, who tend to judge the dots in one of two ways consistently. Either they usually think that the number of dots is more than the reference number, or they usually thinks that it is less. We informed participants that they would be assigned to a "more-estimators" or "less-estimators" group, depending on which response they selected more often in a series of yellow dot trials (13 in total). They then assigned points to a member of the "more" as well as of the "less-estimators" group via the 12 matrices. The question we presented with each matrix was "How many points will you offer in this trial to…", followed by "… member <Y> of the 'less-estimators group'?" and "…member <Z> of the 'more-estimators group'?" (with <Y> and <Z> indicating a random number). Below the response matrix, the following sentence was presented: "You are member 63 of the <more/less>-estimators group. Think carefully."

After completion of the group block, the participants answered three 7-item Likert scale questions measuring Opinionatedness (averaged responses on: "How firmly do you stand behind your dots estimation responses?", "how strongly do you identify with your dots estimation responses?" and "how convinced are you about your dots estimation responses?"). Another three questions measured their levels of Group Identification (averaged responses on: How firmly do you stand behind your group?", "how strongly do you identify with your group?" and "how convinced are you about your group?"). Participants also completed questionnaires related to the processing of disagreement. These included the Rosenberg Self Esteem Scale [3], the Agreeableness subscale of the HEXACO-Pi-R personality questionnaire [4], the Rejection Sensitivity Questionnaire [5] and the Verbal Aggression subscale of the Buss and Perry Aggression Questionnaire [6]. Because self esteem is thought to be influenced by discriminatory behavior [7,8], this questionnaire was presented before the experimental blocks. All other questionnaires were presented after. We asked open-box questions about their experience of the experiment (e.g., technical issues), what their strategy was for assigning points and on whether they noticed anything in particular or wanted to tell us something. Finally, we revealed the deception and asked whether they suspected the other participant(s) not to exist. We assigned them the maximum award of (£0.5) on top of the payment that followed the standards of remuneration at Prolific.

In Exp. 1-3a, we put forward two confirmatory data-analytical approaches in our preregistration: First, we hypothesized that that we would find evidence for the use of discriminatory strategies against a single person in the individual version of the seminal paradigm. Second, we preregistered that we would quantify the strength of the 3 'individual' versus 'group' discrimination pull scores. No specific predictions were made about the directions of these latter potential effects as the group block may be subject to order effects (see discussion). We also put forward one explorative approach that did not involve confirmatory analyses: we proposed to explore via bivariate correlations the relationship of discriminatory tendencies with the aforementioned question(naire)s that measured individual differences. In Exp. 3b, we put forwards the same two confirmatory data-analytical approaches in our preregistration, but this experiment was meant to assess the difference in group versus individual pull scores between-subjects. This experiment did not include any questionnaires.

Experiment 2: Painting Preferences – Within-Subjects. Experiment 2 followed the same instructions, design, hypotheses and procedure as in Experiment 1 except for the following differences: We told the participant that they would have to judge the beauty of artworks and select in each trial which one of two depicted paintings they preferred the most (one shown on the right-hand side and one on the left-hand side). We informed them that raters, on average, were equally likely to choose any of the two paintings. They would assign points to a single individual that had completed the exact same trials already in the past, whose data we had already recorded. In the individual block, each of the trials depicted a painting of Bracques (left) and one of Picasso (right) of roughly the same size. The painters' names were not mentioned to participants. The following question was presented above each response matrix: "How many points will you offer in this trial if the person has chosen…", followed by "…the left painting?" on one row of the matrix and "…the right painting?" on the other. Below the response matrix, the following sentence was presented: "You preferred the <right/left> painting. Think carefully." In the group block, participants were told that they would now choose between art works of the painters Klee and Kandinsky. They were told that people usually tend to prefer one of these painters consistently, and that this allowed us to divide participants into a "Klee group" and a "Kandinsky group". Participants then made their choice between 13 sets of paintings of which there was one of Klee and one of Kandinsky. They were assigned to one of the groups depending on which of the painters they had preferred more across the 13 trials. Afterwards, in the money allocation section of the group block, the question accompanying the group matrices was: "How many points will you offer in this trial to…", followed by "… member <X> of the 'Klee group'?" and "…member <Y> of the 'Kandinsky group'?" (with <X> and <Y> indicating a random number). Below the response matrix we presented: "You are member 63 of the <Klee/Kandinsky>-group. Think carefully."

Experiment 3a: Coin Tosses – Within-Subjects. Experiment 3a also implemented the same general experimental procedures and hypotheses as Experiments 1, except for the following differences. We informed participants that they would repeatedly receive from the computer a head or a tail coin toss of a 50-cent piece, as visualized by a picture. In the individual block, each of the trials depicted either heads or tails of an Australian 50-cent piece. Participants were explicitly told that the outcomes of these coin tosses would be random. They were told that the other person had in the past also received a coin toss in each trial. Each matrix was accompanied by the question: "How many points will you offer in this trial if the person had an…", followed by "…Australian tail as a coin toss?" on one row of the matrix and "…Australian head as a coin toss?" on the other. Below the response matrix, the following sentence was presented: "The coin toss for you was: Australian <head/tail>. Think carefully." In the group block, we used a New Zealand 50-cent piece. We assigned participants to an New Zealand head or a New Zealand tail group, depending which one they received the most after 13 random coin tosses. The matrices here asked "How many points will you offer in this trial to…", followed by "…member <X> of the New Zealand head group?" and "…member <Y> of the New Zealand tail group?" (with <X> and <Y> indicating a random number). Under each matrix was noted: "You are member 63 of the New Zealand <head/tail> group. Think carefully." After the experiment, we used the Group Identification items, Rosenberg Self-Esteem Scale and the Tolerance of Difference questionnaire [9,10].

Experiment 3b: Coin Tosses – Between-Subjects. In this experiment, we re-ran Experiment 3a with the group and the individual tasks manipulated in a between-subjects rather than a within-subjects way. As such, we tested 2 groups of 238 participants each. Here, the coin that was flipped by the computer was always an Australian 50-cent piece, to avoid introducing differences between the groups. Apart from the usual demographical questions, we did not collect questionnaire data in this experiment.

Analyses and Results

Basic Calculation and Computation of Discriminatory Pull Scores. Tajfel's matrices measure the relative strength of an individual's use of a variety of strategies to allocate points. The matrices can register the strength of the two main discriminatory strategies: maximal differentiation (MD) and maximum ingroup profit (MIP). Matrices where these 2 strategies are simultaneously assessed measure the strength of the combined strategy: Favoritism (FAV). The matrices can also capture prosocial strategies: parity (P; an equal division between the two options) and maximum joint profit (MJP; maximising the amount of points assigned). We set out to assess the strength of discriminatory strategies relative to prosocial tendencies. We computed 3 discriminatory pull scores: FAV (MIP + MD) versus MJP, MD versus MIP + MJP, and FAV (MIP + MD) versus P. We employed the computational procedures to determine pull scores described by Bourhis and colleagues. For each of the 6 types of matrices, we averaged the answers on the original and the flipped matrix. Experiments 1-3a involved a 2 (individual/group) x 3 (pull scores) within-subjects design. Experiment 3b involved a 2 (individual/group) x 3 (pull scores) between-within subjects design.

Preregistered Confirmatory Analyses. To evaluate the existence of individual tendencies of unequal resource division and their strength relative to group tendencies, we used a multilevel Bayesian estimation approach to regression in the R programming language [11]. We built models incrementally towards a full model that includes all fixed and varying effects that the design permits [12]. We fit the models using Gaussian distributions via the brms R package [13].

In Experiments 1-3a, to assess whether participants show discriminatory tendencies in both tasks, we used an index coding approach to compare all 6 pull scores and evaluate their posterior distributions compared to zero. Our main focus with this analysis was on the 3 pull scores of the individual block as we predicted these to be different from 0 (which corresponds to no discrimination). The full model (bi2) for each of the experiments looked like this: formula = bf(pull ~ 0 + condition + (0 + condition | pid)), where 'pull' = pull scores, condition = 3 matrix types each for 2 task conditions (individual and group), and pid = participant ID. For Experiment 3b, which used a between-participant design, the model formula was identical, but we fit two regression models, one for each participant group in our design. For the results, see figures in the main text and suppl. tables 2, 4, 6.

To examine the strength of the pull scores across the two blocks in Experiments 1-3a, we built a separate second Bayesian regression model. We used a deviation coding approach, which resembles the structure of an ANOVA where factors sum to zero. Then we compared average effects of task (individual vs. group), matrixtype (mat1 = FAV vs. MJP, mat2 = MD vs. MIP/MJP, and mat3 = FAV vs. P) and their interaction. The full model (bd5.1) for each of the experiments looked like this: formula = bf(pull ~ 1 + task * mat1 + task * mat2 + task * mat3 + (1 + task * mat1 + task * mat2 + task * mat3 | pid)). For the results, see figures in the main text and suppl. tables 3, 5, 7.

In Experiment 3b, we followed the same logic. We compared the effects of matrixtype (FAV vs. MJP, MD vs. MIP/MJP, and FAV vs. P) as a function of task (individual vs. group). The regression formula for the full model looked like this: formula = bf(pull ~ 1 + task * mat2 + task * mat3 + (1 + mat2 + mat3 | pid)) where 'pull' = pull scores, task = 'individual' vs 'group', mat2 = matrix 2 vs matrix 1, mat3 = matrix 3 vs matrix 1, and pid = participant ID.

For the results, see figures in the main text and suppl. tables 8 and 9. Two differences are noteworthy for Exp. 3b compared to Exps. 1-3a. First, as previously mentioned, task cannot vary by participant, which means it cannot be set as a varying effect. Second, we removed a matrix term, as we realised that the ANOVA structure could be specified in a simpler manner (see here for details: https://debruine.github.io/faux/articles/contrasts.html#anova).

Preregistered Exploratory Analyses: Correlations. We used bivariate correlations to explore the relationship between discriminatory behaviour in the individual and group blocks of Experiments 1-3a where appropriate with the measures of self-esteem, opinionatedness, group identification, agreeableness, verbal aggression, rejection sensitivity, and tolerance for difference. As can be seen in suppl. Tables 10 to 12, all correlations in the dots estimations, painting preferences and coin toss experiments were of a small size of between -0.15 and 0.15, except for medium-to-large sized positive correlations of Group Identification and some of the group pull scores.

Preregistered Exploratory Analyses: Frequentist Statistics. For reasons of comparison, we include here frequentist (non-Bayesian) statistical approaches that are conventional in original minimal group studies[1]. We implemented a paired-samples Wilcoxon signed rank test in R on the participants' answers in the together versus opposed version (for more information, see [1]) of each of the 3 types of matrix (FAV vs. MJP, MD vs. MIP/MJP, and FAV vs. P) and per task (individual versus group). In Experiment 1 (dots estimations), we observed for the individual task version significant effects in each type of response matrix (FAV vs. MJP: $V = 4140.5$, $p < .000$, Cohen's $d = .52$; MD vs. MIP/MJP: $V = 5998$, $p < .02$, Cohen's $d = .16$; and FAV vs. P: $V = 3081.5$, $p < .000$, Cohen's $d = .56$). The same was true for the group version (FAV vs. MJP: $V = 1645$, $p < .000$, Cohen's $d = .59$; MD vs. MIP/MJP: $V = 2683$, $p < .000$, Cohen's $d = .27$; and FAV vs. P: $V = 1737.5$, $p < .000$, Cohen's $d = .58$). We then implemented a 2 (task) x 3 (matrixtype) ANOVA on the pull scores. This yielded a main effect of matrixtype ($F = 82.02$, $p < 000$) and of task ($F = 6.61$, $p = 0.01$) but no interaction. In Experiment 2 (painting preferences), similar patterns emerged: The individual task version showed significant effects in each type of response matrix (FAV vs. MJP: $V = 1310$, $p < .000$, Cohen's $d = .68$; MD vs. MIP/MJP: $V = 2261$, $p < .000$, Cohen's $d = .44$; and FAV vs. P: $V = 1468$, $p < .000$, Cohen's $d = .65$), as did the group tasks version (FAV vs. MJP: $V = 511.5$, $p < .000$, Cohen's $d = .70$; MD vs. MIP/MJP: $V = 783.5$, $p < .000$, Cohen's $d = .53$; and FAV vs. P: $V = 619$, $p < .000$, Cohen's $d = .69$). The 2 x 3 ANOVA yielded a main effect of matrixtype ($F = 75.47$, $p < .000$) and of task ($F = 17.46$, $p <.000$), but not interaction. In Experiment 3a (coin tosses - within), the individual task version showed significant effects in each type of response matrix (FAV vs. MJP: $V = 2589.5$, $p < .000$, Cohen's $d = .47$; MD vs. MIP/MJP: $V = 2412.5$, $p < .05$, Cohen's $d = .14$; and FAV vs. P: $V = 2805$, $p < .000$, Cohen's $d = .48$), as did the group task version (FAV vs. MJP: $V = 661.5$, $p < .000$, Cohen's $d = .58$; MD vs. MIP/MJP: $V = 1427$, $p < .000$, Cohen's $d = .32$; and FAV vs. P: $V = 861$, $p < .000$, Cohen's $d = .55$). The 2 x 3 ANOVA yielded a main effect of matrixtype ($F = 80.07$, $p < .000$) and of task ($F = 4.25$, $p < .05$), but no interaction. In Experiment 3b (coin tosses - between), the individual task version showed significant effects in all response matrices expect for one where the effect was marginally significant (FAV vs. MJP: $V = 2958.5$, $p < .000$, Cohen's $d = .46$; MD vs. MIP/MJP: $V = 2110.5$, $p = .08$, Cohen's $d = .11$; and FAV vs. P: $V = 2631.5$, $p < .000$, Cohen's $d = .47$), and in all types of the group task version (FAV vs. MJP: $V = 1540$, $p < .000$, Cohen's $d = .66$; MD vs. MIP/MJP: $V = 1714$, $p < .000$, Cohen's $d = .50$; and FAV vs. P: $V = 1486.5$, $p < .000$, Cohen's $d = .66$). The 2 x 3 ANOVA yielded a main effect of matrixtype ($F = 1.85$, $p < .000$) and of task ($F = 24.5$, $p < .000$), but no interaction.

Like the Bayesian results, these findings yielded strong evidence for the existence of discriminatory behavior against a single individual that demonstrates a different versus the same outcome in dots estimations, painting preferences or coin tosses – with often large effect sizes (Cohen's d > .050).

PART 2: Experiments Based on Cognitive Neuroscience's Congruency Design

Participants

We conducted Experiments 4, 5 and 6 with a sample of 75 participants each. This sample size was determined using the effect size data in our first 3 discrimination experiments and largely following our earlier sample size rationale while increasing the power. The congruency experiments do not differentiate between an MD and MIP strategy. We therefore relied on the effect sizes of the two pull scores in the individual block that measured the pull of a discriminatory strategy versus a prosocial strategy (i.e., for FAV (MD + MIP) versus MJP; and FAV (MD + MIP) versus P), rather than versus one another (i.e. MD versus MIP/MJP). The minimum effect size for these two pull scores was $d = 0.51$ to $d = 0.81$ across Experiments 1-3a. We thus calculated the participant number via an a-priori power analysis in the program G*Power, based on a matched-pairs t-test with parameters: one-tailed, effect size of 0.51, alpha = 0.05, power is 0.95, normal parent distribution. This yielded a sample size of $N = 44$. Because we could afford to collect slightly more data and wanted to maximize the data to collect, we set a target of $N = 75$ in this experiment. We used the same eligibility and other participant criteria as in our first three experiments. We replaced participant data when they had experienced technical issues (Exp. 4: 1 dataset, Exp. 5: 0 datasets, and Exp. 6: 1 dataset).

Design

These experiments adapted the individual dots estimations, painting preferences and coin toss trials into a social congruency design. The trials of Experiments 4 to 6 had the same initial set-up as those of Experiments 1 to 3ab respectively: the participant needed to determine whether a cloud of dots is more or less than a reference number, they chose which painting they preferred out of two paintings, or they received an Australian heads or tails coin flip. The participant also assigned points to the other, but not vice versa. We hence ruled out any influence of social reciprocity or self-interest in the participants' point allocation measures [14,15] because the other person did not assign points in turn. We describe the other differences with Experiments 1 to 3ab hereafter.

Experiment 4: Dots Estimations. Instead of using the matrices dependent measures, the participant was told that they would be given a maximum number of points to assign to the other person (e.g., 20 points), and to use their keyboard to enter the amount of points they want to give (e.g., 10 points). After estimating the amount of dots as more or less than the reference number, they viewed a fixation cross for 500ms. They were then given feedback for a duration of 1500ms about another person's choice: "The other participant thought that the amount of points was <more/less> than the reference number". In half of the trials, this feedback was identical to the participant's own decision (congruent trial), and in the other half, the opposite (incongruent trial). Then, we presented "you can now assign to this person a maximum of <10/20/30/40/50/60/70/80/90/100> points. Type your response. Think carefully". After this, the next trial would start. Each maximum number was shown once per condition, for a total of 550 points per condition. We thus presented 10 trials for each condition in randomized order. We computed for each trial the percentage assigned points out of the respective maximum amount of points for that trial. We preregistered the following hypothesis: that the percentage of assigned points will be larger when the other individual chose the same dots estimation as the participant (congruent condition) compared to a different dots estimation (incongruent condition).

Experiment 5: Painting Preferences. This experiment used a series of 10 sets of two Klee and Kandinsky paintings (unnamed) in each condition. After choosing one of the two paintings in each trial, the participant was given congruent or incongruent information about the choice of the other person: “The other participant preferred the <left/right> painting”.

Experiment 6: Coin Tosses. This experiment used the Australian head and tail stimuli of the previous coin toss experiment (Exp. 2). After receiving their own coin toss outcome, the congruent or incongruent feedback noted: “The other participant received the Australian <head/tail> outcome”.

Analyses and Results

Preregistered Confirmatory Analyses. In Experiments 4 to 6, we used the same general approach to data analysis that we used in Experiments 1 to 3. To assess whether participants show discriminatory tendencies in the face of disagreement about paintings, we included a categorical predictor to estimate the average effects of condition (congruent vs. incongruent). Levels of condition were specified using a deviation coding approach (e.g., congruent = 0.5, incongruent = -0.5).

The full model (b3) looked like this: formula = bf(percentage ~ 1 + condition + (1 + condition | pid)). Note: percentage = percentage of assigned points; condition = congruent vs. incongruent, and; pid = participant ID. To support our primary hypothesis, we expected the posterior distribution for the average effect of condition to be positive such that the percentage of points assigned to the other person is higher in the congruent than incongruent condition. For the results, see figures in the main text and suppl. table 11.

Frequentist Analyses (not preregistered). For reasons of comparison, we add frequentist statistics for Experiments 4 to 6 as well. The Wilcoxon paired samples test on the percentages of money assigned in the congruent versus incongruent conditions in each experiment yielded highly significant results and large effect sizes (Exp. 1: $V = 1959.5$, $p <.000$, Cohen's $d = .75$; Exp. 2: $V = 2112$, $p <.000$, Cohen's $d = .78$; and Exp. 3: $V = 1400.5$, $p <.000$, Cohen's $d = .52$). Like in the Bayesian results, strong evidence is obtained for the existence of discrimination against a single individual that demonstrates a different versus the same outcome in dots estimations, painting preferences and coin tosses.

Tables

Table S1: Demographic characteristics in Experiments 1 to 6.

		Exp. 1: (n = 238)	Exp. 2: (n=238)	Exp. 3a: (n=238)	Exp. 3b: Individual task (n=238)	Exp.3b: Group task (n=238)	Exp. 4: (n=75)	Exp. 5: (n=75)	Exp. 6: (n=75)
Gender	Male	50.84%	49.16%	48.74%	57.56%	56.72%	48.0%	46.67%	49.33%
	Female	49.16%	50.42%	50.84%	41.60%	42.86%	48.0%	52.00%	50.67%
	Other	0.00%	0.00%	0.00%	0.00%	0.00%	1.33%	0.00%	0.00%
	Prefer not to say	0.00%	0.42%	0.42%	0.42%	0.00%	0.00%	1.33%	0.00%
Education	Elementary school	0.00%	0.00%	0.00%	0.42%	0.00%	0.00%	0.00%	0.00%
	High school	17.65%	13.03%	13.45%	11.76%	17.65%	12.00%	16.00%	14.67%
	Some college or further	11.76%	14.29%	13.45%	13.87%	11.76%	8.00%	12.00%	16.00%
	Trade or vocational school	6.30%	5.04%	6.72%	3.36%	6.30%	4.00%	5.33%	1.33%
	Post-secondary certificate or diploma	6.30%	4.20%	1.26%	6.72%	6.30%	6.67%	4.00%	4.00%
	Bachelor's degree	37.82%	42.44%	43.70%	42.02%	37.82%	44.00%	41.33%	42.67%
	Master's degree	17.23%	16.39%	17.65%	17.65%	17.23%	20.00%	12.00%	20.00%
	Doctorate	2.94%	4.62%	3.78%	4.20%	2.94%	4.00%	9.33%	1.33%
Age	Mean (years)	43.67	42.18	39.68	43.13	43.67	45.55	43.99	41.13
	Stdev (years)	13.50	13.35	11.58	13.32	13.50	14.22	13.30	13.33
Handedness	Right	84.87%	84.03%	84.87%	85.29%	84.87%	88.00%	89.33%	93.33%
	Left	10.92%	13.03%	14.29%	12.61%	10.92%	10.67%	8.00%	6.67%
	Ambidextrous	4.20%	2.94%	0.84%	2.10%	4.20%	0.00%	2.67%	0.00%

Table S2: Model bi2 (index) fixed effects in original units for Exp. 1 (dots estimations – within subjects).

Model bi2 fixed effects

task	*matrix_name*	*value*	*.lower*	*.upper*
individual	FAVvsMJP	2.38	1.89	2.85
individual	MDvsMIP/MJP	0.43	0.09	0.78
individual	FAVvsP	2.81	2.30	3.31
group	FAVvsMJP	3.27	2.74	3.79
group	MDvsMIP/MJP	0.97	0.56	1.37
group	FAVvsP	3.17	2.65	3.69

Table S3: Model bi5 fixed effects (factorial) in original units for Exp. 1 (dots estimations – within subjects).

Model bd5.1 fixed effects

term	*value*	*.lower*	*.upper*
intercept	2.43	2.07	2.78
task	0.66	0.19	1.13
FAVvsMJP	0.73	-0.43	1.88
MDvsMIP/MJP	-1.64	-2.83	-0.47
FAVvsP	0.91	-0.24	2.05
task*FAVvsMJP	0.33	-0.83	1.48
task*MDvsMIP/MJP	-0.11	-1.27	1.08
task*FAVvsP	-0.22	-1.38	0.94

Table S4: Model bi2 fixed effects (index) in original units for Exp. 2 (paintings preferences – within subjects).

Model bi2 fixed effects

task	*matrix_name*	*value*	*.lower*	*.upper*
individual	FAVvsMJP	3.16	2.69	3.63
individual	MDvsMIP/MJP	1.40	1.03	1.77
individual	FAVvsP	3.30	2.79	3.78
group	FAVvsMJP	4.16	3.63	4.66
group	MDvsMIP/MJP	1.97	1.53	2.39
group	FAVvsP	4.33	3.80	4.84

Table S5: Model bi5 fixed effects (factorial) in original units for Exp. 2 (painting preferences – within subjects).

Model bd5.1 fixed effects

term	*value*	*.lower*	*.upper*
intercept	3.40	3.02	3.76
task	0.95	0.52	1.37
FAVvsMJP	0.66	-0.48	1.82
MDvsMIP/MJP	-1.53	-2.69	-0.35
FAVvsP	0.84	-0.31	2.00
task*FAVvsMJP	0.14	-1.04	1.33
task*MDvsMIP/MJP	-0.31	-1.51	0.90
task*FAVvsP	0.16	-1.02	1.37

Table S6: Model bi2 fixed effects (index) in original units for Exp. 3a (coin tosses - within-subjects).

Model bi2 fixed effects

task	*matrix_name*	*value*	*.lower*	*.upper*
individual	FAVvsMJP	2.53	1.98	3.10
individual	MDvsMIP/MJP	0.35	0.06	0.65
individual	FAVvsP	2.60	2.04	3.17
group	FAVvsMJP	3.12	2.58	3.65
group	MDvsMIP/MJP	0.97	0.61	1.33
group	FAVvsP	3.20	2.65	3.74

Table S7: Model b5.1 (factorial) fixed effects in original units for Exp. 3a (coin tosses – within-subjects).

Model bd5.1 fixed effects

term	*value*	*.lower*	*.upper*
intercept	2.37	2.02	2.72
task	0.63	0.12	1.13
FAVvsMJP	0.76	-0.36	1.93
MDvsMIP/MJP	-1.63	-2.80	-0.47
FAVvsP	0.85	-0.28	2.02
task*FAVvsMJP	-0.03	-1.23	1.18
task*MDvsMIP/MJP	0.05	-1.18	1.31
task*FAVvsP	-0.03	-1.22	1.18

Table S8: Model bi2 fixed effects (index) in original units for Exp. 3b (coin tosses - between-subjects).

Model bi2 fixed effects

task	*matrix_name*	*value*	*.lower*	*.upper*
individual	FAVvsMJP	2.6	2.06	3.16
individual	MDvsMIP/MJP	0.33	0.03	0.64
individual	FAVvsP	2.58	2.03	3.14
group	FAVvsMJP	4.06	3.51	4.6
group	MDvsMIP/MJP	2.1	1.65	2.57
group	FAVvsP	4.17	3.61	4.71

Table S9: Model b5.1 (factorial) fixed effects in original units for Exp. 3b (coin tosses – between-subjects).

Model bd5.1 fixed effects

term	*value*	*.lower*	*.upper*
intercept	2.92	2.58	3.26
task	1.70	1.06	2.33
MDvsMIP/MJP	-2.29	-2.72	-1.83
FAVvsP	0.07	-0.23	0.39
task*MDvsMIP/MJP	0.40	-0.38	1.21
task*FAVvsP	-0.09	-0.5	0.67

Table S10. Bivariate correlations in Exp. 1 (dots estimations).

Bivariate correlations in Exp. 1							
		Opinion	*GroupId*	*SelfEst*	*Agreeab*	*RejectSens*	*VerbAg*
Individual	*FAVvsMJP*	0	0.08	0.01	0.07	-0.11	0.04
	MDvsMIP/MJP	0.03	0.01	0.04	0.11	-0.01	-0.08
	FAVvsP	0.01	0.13	-0.1	0.09	0	-0.04
Group	*FAVvsMJP*	0.15	0.29	-0.15	-0.01	0.02	0.06
	MDvsMIP/MJP	0.04	0.13	-0.08	-0.01	-0.08	0.08
	FAVvsP	0.11	0.31	-0.06	0.08	-0.07	0.01

Note. Correlations between the relevant questionnaires and the pull scores of both the group and the individual block. Abbreviations for questionnaires from left to right: Opinionatedness, Group Identification, Self Esteem, Agreeableness, Rejection Sensitivity, Verbal Aggression.

Table S11. Bivariate correlations in Exp. 2 (painting preferences).

Bivariate correlations in Exp. 2							
		Opinion	*GroupId*	*SelfEst*	*Agreeab*	*RejectSens*	*VerbAg*
Individual	*FAVvsMJP*	0.01	0.15	0.02	0.04	-0.15	0.01
	MDvsMIP/MJP	0.02	0.14	-0.05	-0.01	0.02	0.04
	FAVvsP	-0.06	0.08	0.02	0.07	-0.12	0
Group	*FAVvsMJP*	0.04	0.29	0.02	-0.02	-0.05	0.01
	MDvsMIP/MJP	0.09	0.22	0	-0.01	-0.01	0.09
	FAVvsP	-0.01	0.24	0.01	0.03	-0.04	0

Note. Correlations between the relevant questionnaires and the pull scores of both the group and the individual block. Abbreviations for questionnaires from left to right: Opinionatedness, Group Identification, Self Esteem, Agreeableness, Rejection Sensitivity, Verbal Aggression.

Table S12. Bivariate correlations in Exp. 3 (coin tosses).

Bivariate correlations in Exp. 3				
		GroupId	*SelfEst*	*TolDiff*
Individual	*FAVvsMJP*	0.11	-0.02	0.07
	MDvsMIP/MJP	0.03	-0.07	0.04
	FAVvsP	0.22	0.07	0.12
Group	*FAVvsMJP*	0.52	0.03	0.05
	MDvsMIP/MJP	0.19	-0.05	-0.05
	FAVvsP	0.49	-0.04	0.04

Note. Correlations between the relevant questionnaires and the pull scores of both the group and the individual block. Abbreviations for questionnaires from left to right: Group Identification, Self Esteem, Tolerance for Difference.

Table S13: Model b3 fixed effects in original units for Exp. 4 to 6 (congruency versions of dots estimations, paintings preferences and coin tosses respectively).

Model b3 fixed effects in original units

term	*exp*	*value*	*.lower*	*.upper*
intercept	exp4_dots	71.19	65.62	76.83
intercept	exp5_paintings	64.79	58.38	70.11
intercept	exp6_coins	64.17	58.73	69.58
condition	exp4_dots	23.71	18.19	29.37
condition	exp5_paintings	24.69	18.18	30.72
condition	exp6_coins	17.87	11.56	24.14

Figures

Figure S1. Examples of matrices used in Experiments 1 to 3ab. A. Individual matrices. B. Group matrices.

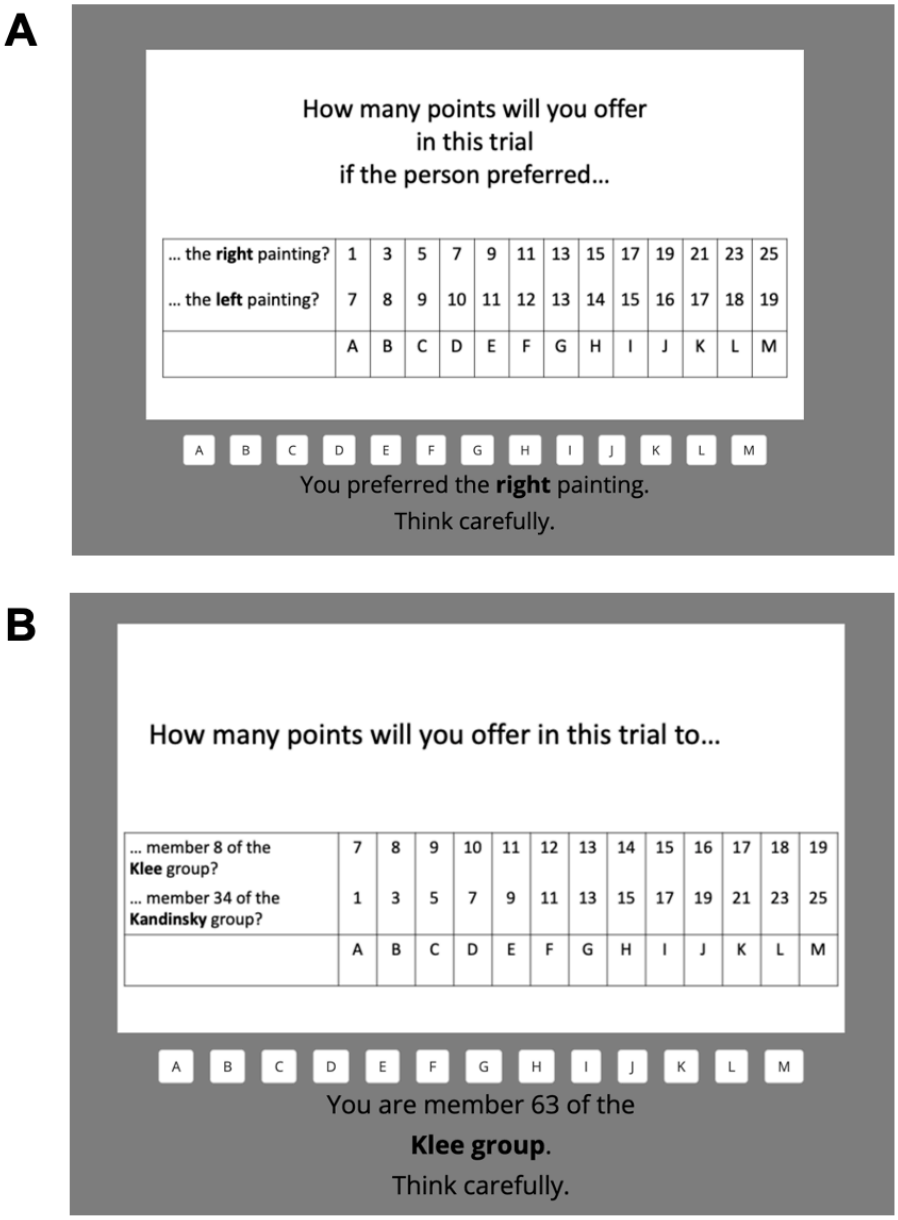


Figure S2. Designs of the Experiments 1 to 3ab, individual versions.

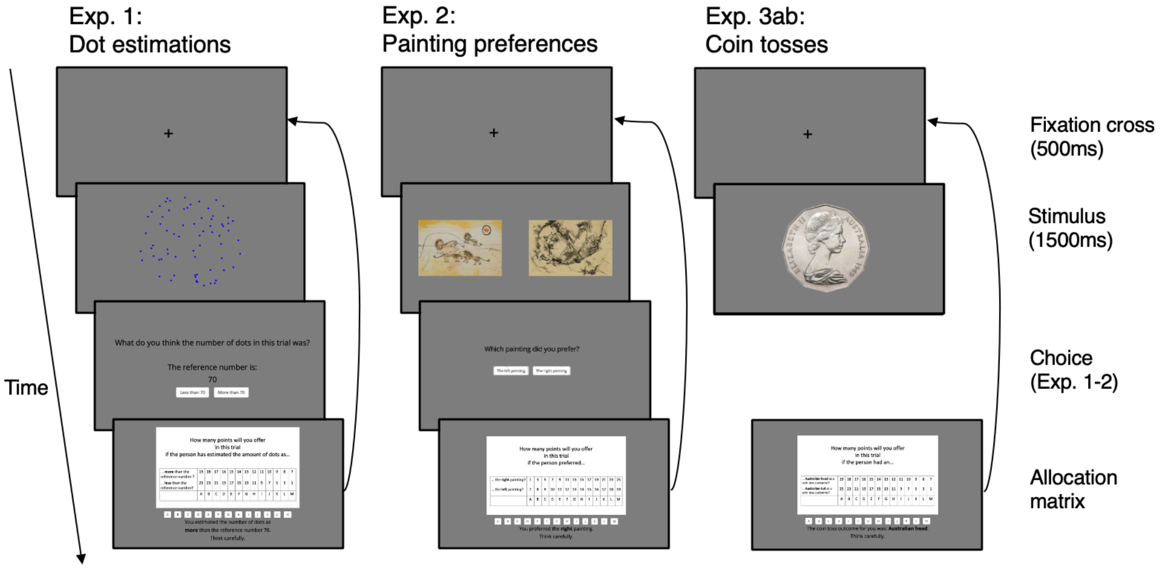


Figure S3. Designs of the Experiments 1 to 3ab, group versions.

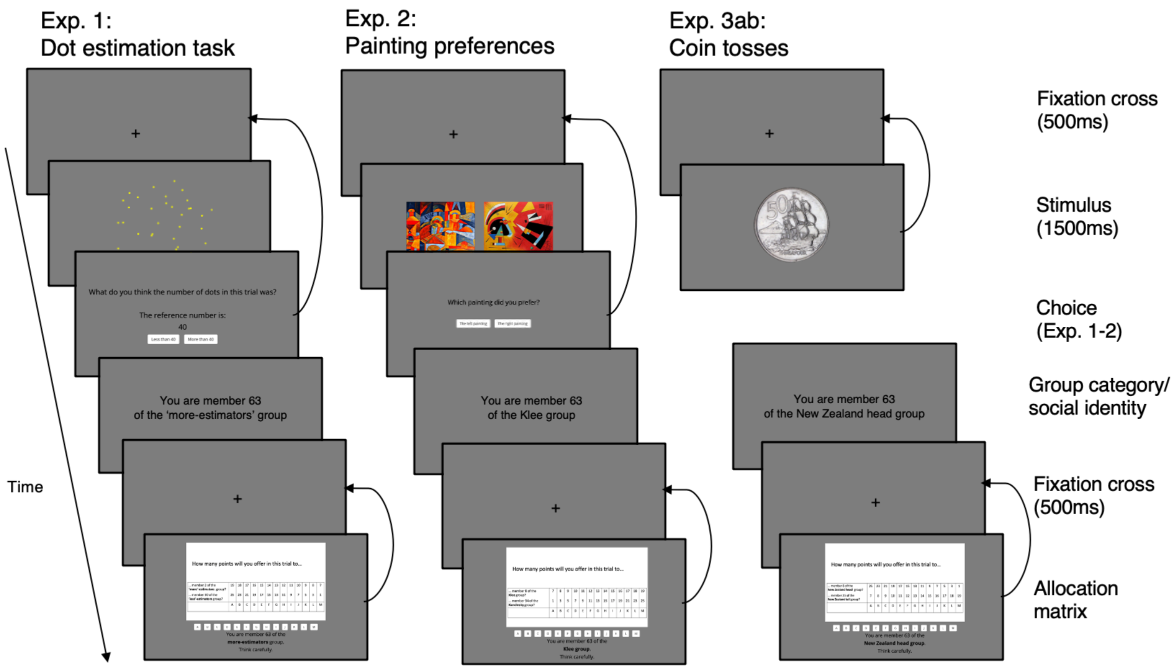


Figure S4. Designs of the Experiments 4 to 6.

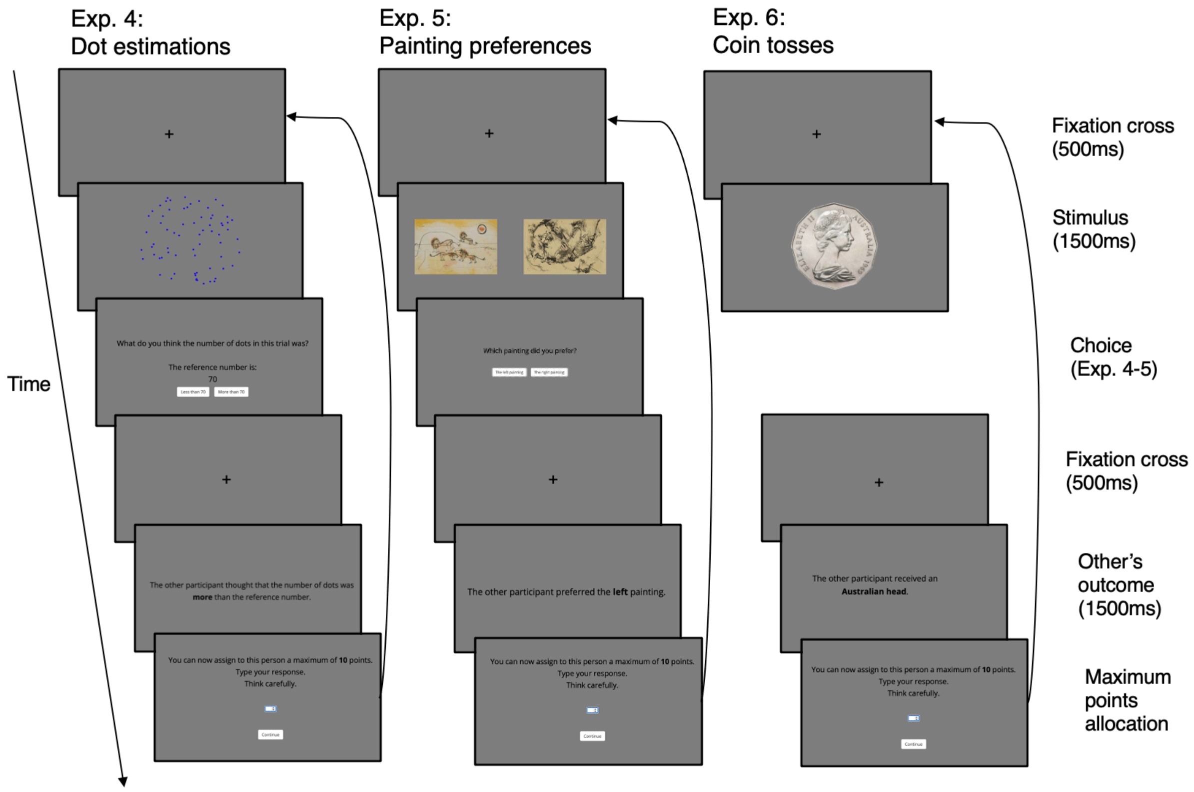
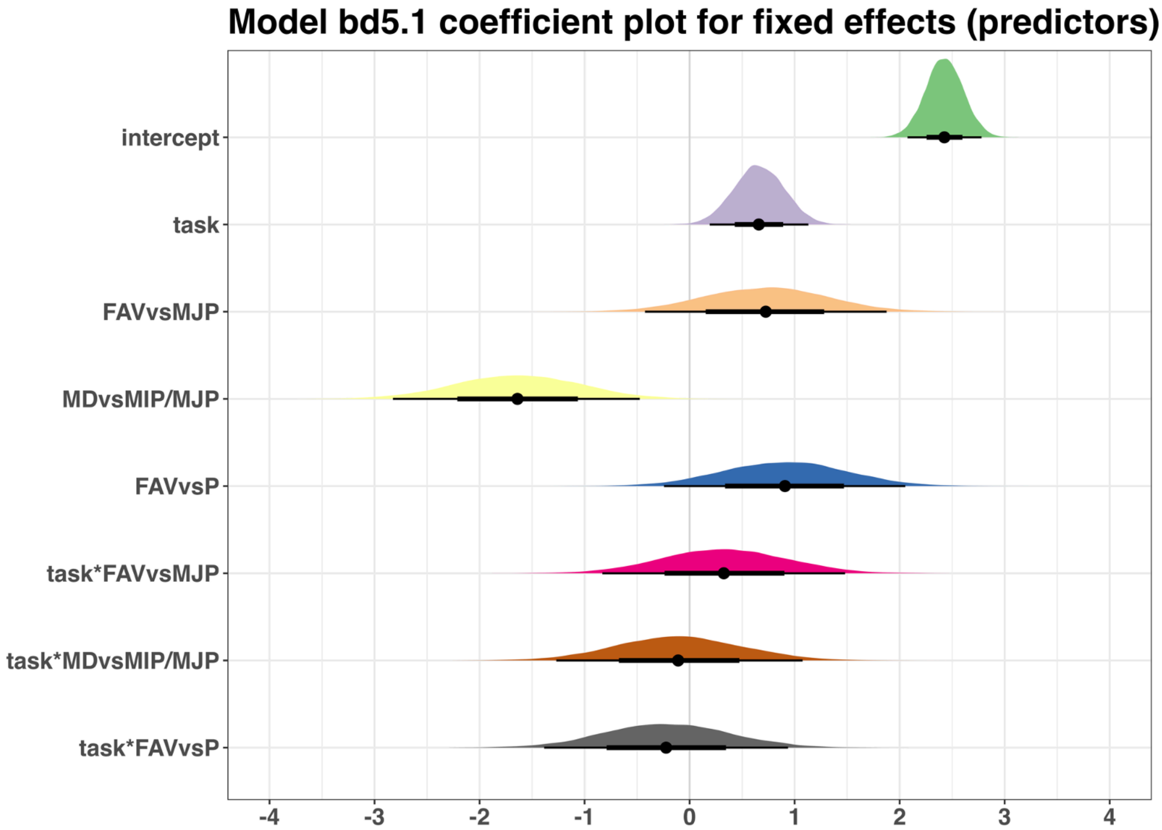
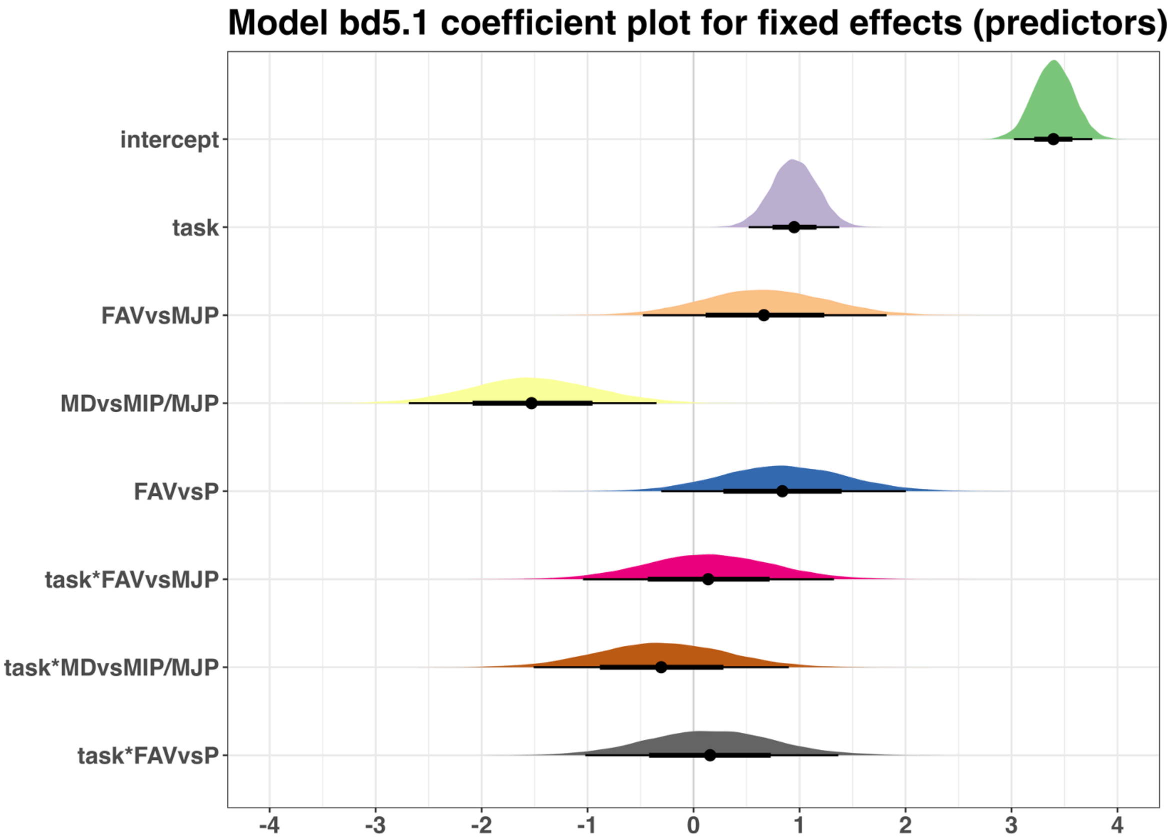


Figure S5: Bayesian model (bd5.1) to test for the effect of block (task: individual versus group) in Experiment 1 (dots estimations).



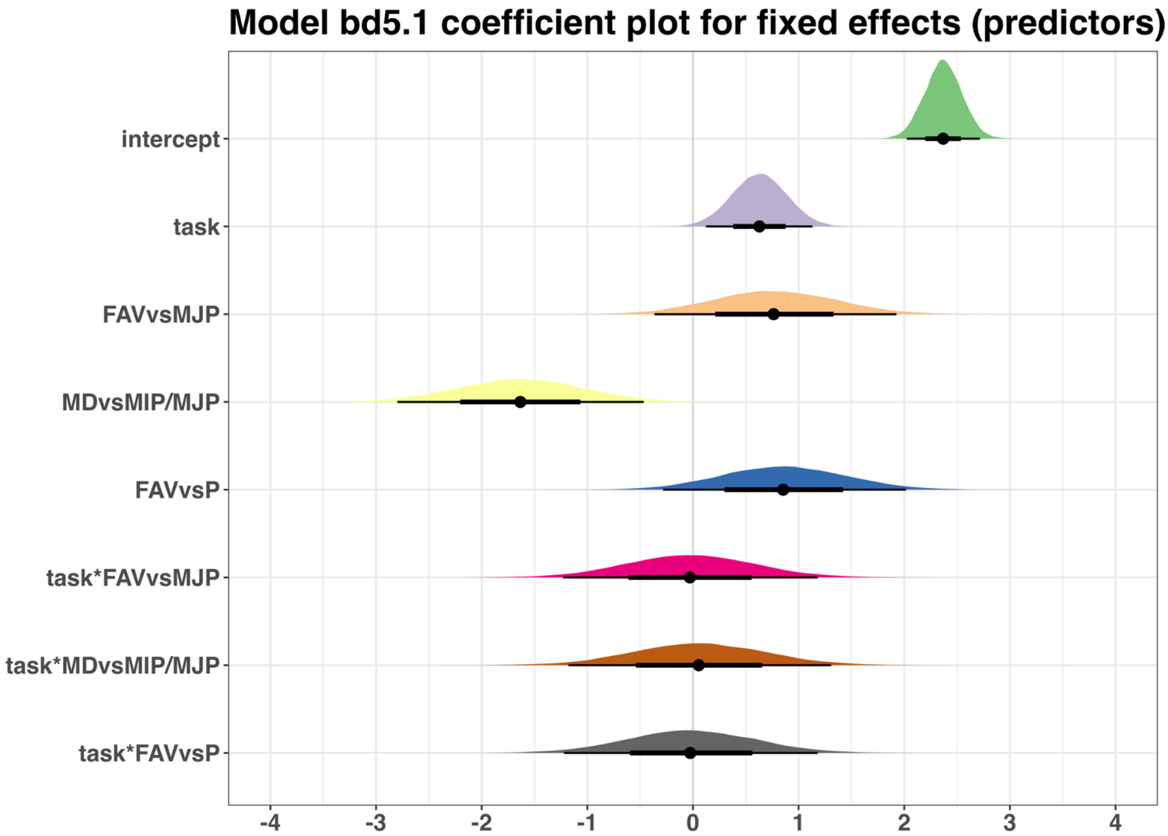
Note. Error bars denote the 66 and 95% quantile intervals in the coefficient plot. The coefficient for the fixed effect of task (block) is above 0, suggesting that the pull scores in the group task were somewhat larger than those in the individual task.

Figure S6: Bayesian model (bd5.1) to test for the effect of block (task: individual versus group) in Experiment 2 (paintings preferences).



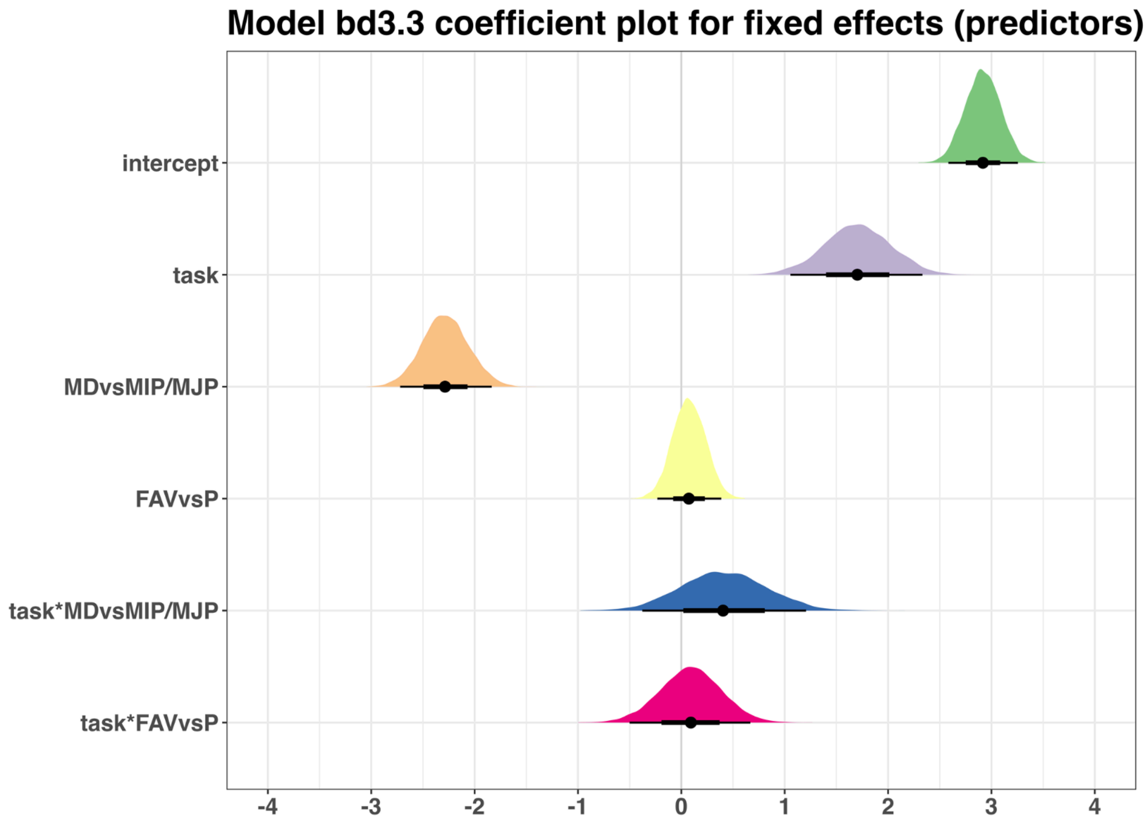
Note. Error bars denote the 66 and 95% quantile intervals in the coefficient plot. The coefficient for the fixed effect of task (block) is slightly above 0, suggesting that the pull scores in the group task were somewhat larger than those in the individual task.

Figure S7: Bayesian model (bd5.1) to test for the effect of block (task: individual versus group) in Experiment 3a (coin tosses – within-subjects).



Note. The coefficient for the fixed effect of task (block) is slightly above 0, suggesting that the pull scores in the group task were somewhat larger than those in the individual task. Error bars denote the 66 and 95% quantile intervals in the coefficient plot.

Figure S8: Bayesian model (bd5.1) to test for the between-subjects effect (task: individual versus group) in Experiment 3b (coin tosses – between-subjects).



Note. The coefficient for the fixed effect of task is above 0, suggesting that the pull scores in the group task were larger than those in the individual task. Error bars denote the 66 and 95% quantile intervals in the coefficient plot.

SI References

1. Bourhis, R. Y., Sachdev, I. & Gagnon, A. Intergroup Research With the Tajfel Matrices. *Psychol. Prejud. Ontario Symp.* 209–228 (1994).
2. Munafò, M. R. *et al.* A manifesto for reproducible science. *Nature Human Behaviour* vol. 1 (2017).
3. Rosenberg, M. *Society and the adolescent self-image*. (Princeton University Press, 1965).
4. Lee, K. & Ashton, M. C. The HEXACO-PI-R: A measure of the six major dimensions of personality. *Twin Res. Hum. Genet.* **10**, 799–805 (2009).
5. Downey, G. & Feldman, S. I. Implications of rejection sensitivity for intimate relationships. *J. Pers. Soc. Psychol.* **70**, 1327–1343 (1996).
6. Buss, A. H. & Perry, M. The aggression questionnaire. *J. Pers. Soc. Psychol.* **63**, 452–459 (1992).
7. Lemyre, L. & Smith, P. M. Intergroup discrimination and self-esteem in the minimal group paradigm. *J. Pers. Soc. Psychol.* **49**, 660–670 (1985).
8. Chin, M. G. & McClintock, C. G. The effects of intergroup discrimination and social values on level of self-esteem in the minimal group paradigm. *Eur. J. Soc. Psychol.* **23**, 63–75 (1993).
9. Hjerm, M., Eger, M. A., Bohman, A. & Fors Connolly, F. A New Approach to the Study of Tolerance: Conceptualizing and Measuring Acceptance, Respect, and Appreciation of Difference. *Soc. Indic. Res.* **147**, 897–919 (2020).
10. Cuadrado, I., Ordóñez-Carrasco, J. L., López-Rodríguez, L., Vázquez, A. & Brambilla, M. Tolerance towards difference: Adaptation and psychometric properties of the Spanish version of a new measure of tolerance and sex-moderated relations with prejudice. *Int. J. Intercult. Relations* **84**, 220–232 (2021).
11. McElreath, R. *Statistical Rethinking*. *Statistical Rethinking* (CRC Press, 2018). doi:10.1201/9781315372495.
12. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).
13. Bürkner, P. C. brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, (2017).
14. Yamagishi, T. & Kiyonari, T. The Group as the Container of Generalized Reciprocity. *Soc. Psychol. (Gott).* **63**, 116–132 (2000).
15. Rabbie, J. M. & Horwitz, M. Arousal of ingroup-outgroup bias by a chance win or loss. *J. Pers. Soc.*

Psychol. **13**, 269–277 (1969).